To cite this chapter:

Dejonckheere, E., & Mestdagh, M. (2021). On the signal-to-noise ratio in real-life emotional time series. In C. E. Waugh & P. Kuppens (Eds.), *Affect dynamics* (pp. 131-152). Springer. https://doi.org/10.1007/978-3-030-82965-0_7

Chapter 7
On the Signal-to-Noise Ratio in Real-Life Emotional Time Series



Egon Dejonckheere and Merijn Mestdagh

Abstract How emotions unfold in everyday life, naturally interact with each other over time, change in function of different real-life situations or contexts, and how these dynamic patterns relate to person-level characteristics are areas of investigation that are gaining increased interest among emotion researchers. At the same time, however, there is also growing awareness that adequately assessing affect dynamics 'in the wild' may be harder than previously anticipated. A central predicament is that the data quality of current daily life studies is often insufficient to pick up on meaningful regularities in real-life emotion time series, concealing the true dynamical features of people's affective life. In this chapter, we introduce the signal-to-noise ratio (SNR) as a measure of data quality, designed to disentangle the power of people's latent emotional signal from inevitable background noise. Breaking down the SNR into its different constituents, we lay out an overarching framework with various methodological strategies to improve the SNR of real-life affective time series, with the ultimate goal to reliably evaluate the internal and external validity of ecological dynamics of affect. Providing preliminary empirical evidence, we hope that future daily life studies will implement our suggestions, to truly behold the dynamical nature of everyday emotion.

Keywords Affect dynamics · Signal-to-noise ratio · Measurement error · Innovation · Context · Auto-correlation · Measurement burst · Inertia

7.1 Introduction

Investigating how affective states wax and wane in the realm of everyday life is an essential complement to studying emotional responding in standardized, yet often artificial, lab settings. Unparalleled in their ecological validity, daily life methods

E. Dejonckheere (✉) · M. Mestdagh
Faculty of Psychology and Educational Sciences, Leuven, Belgium
e-mail: egon.dejonckheere@kuleuven.be; merijn.mestdagh@kuleuven.be

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

C. E. Waugh, P. Kuppens (eds.), *Affect Dynamics*,
https://doi.org/10.1007/978-3-030-82965-0_7

such as experience sampling (ESM) provide researchers with a unique first-seat insight into the authentic structure and natural dynamical regularities of people's emotional life. Within individuals, for example, ESM researchers may show interest in the circadian rhythms of emotion (e.g., Murray et al., 2009), the ability of an emotional state to predict its future trajectory (e.g., emotional inertia, Cole & Hollenstein, 2018), the concurrent and temporal interplay between various emotions (e.g., the relation between positive [PA] and negative affect [NA], Dejonckheere et al., 2018; emotional augmentation and blunting, Pe & Kuppens, 2012), and natural emotional change as a function of different internal and external processes or stimuli (e.g., emotion regulation, Kalokerinos et al., 2017; encountering negative events, Koval et al., 2015). In turn, between individuals, ESM studies may investigate how these dynamical patterns of emotion are informative for various person-level features or characteristics, such as psychological well-being or maladjustment (e.g., Houben et al., 2015), personality traits (e.g., Bringmann et al., 2016) or emotional intelligence (e.g., Robinson et al., 2020).

Although each of these studies typically require a unique and tailored ESM protocol to validly answer their research questions, a common prerequisite for all their designs is that they accurately describe the real-life patterns of people's emotional life. If ESM time series do not reliably represent the natural ebbs and flows of participants' affective experiences, emotion researchers run the risk of drawing erroneous conclusions about emotions' true dynamic properties, their causes and consequences (i.e., together referred to as the *internal* validity of affect dynamics), but also how differences in these temporal trajectories of emotion may explain individual variation in important person-level variables (i.e., the *external* validity of affect dynamics).

Today, there is growing consensus that an accurate evaluation of the internal and external validity of real-life affect dynamics may be harder than previously thought. Regarding the internal validity, for example, recent research shows that trying to determine a person's current emotional state is less effective when relying on rather complex dynamical models of affect compared to the simplified version of these models (Bulteel et al., 2018). This is remarkable, because more complex models allow ESM researchers to incorporate multiple streams of information (e.g., personal or contextual data), and are therefore thought to be a closer approximation of reality. In contrast, the very basic models merely rely on a person's previous emotions to determine his or her current emotional state, disregarding other potentially useful sources of information relevant for the conception of an emotional response (e.g., Frijda, 1988). Nevertheless, these simple affect dynamic models outperform the complex ones in terms of predictive accuracy (Bulteel et al., 2018), raising the question whether there may be methodological issues associated with ESM that currently conceal the complex (yet true) dynamical features of people's affective life.

Second, also the external validity of affect dynamics was recently questioned in various psychological domains. From common symptom types in mental health research (Dejonckheere et al., 2019a), to the big five traits in the study of personality (Hisler et al., 2020; Kalokerinos et al., 2020; Wendt et al., 2020), to multiple facets of emotional intelligence (MacCann et al., 2020), recent research shows that

more complex metrics of affect dynamics no longer explain individual differences in these person-level outcomes once the predictive power of simple mean levels of PA and NA is taken into account. Again, the observation that the dynamical regularities of people's emotional life play a negligible role in our understanding of their well-being, personality or emotional intelligence seems to be counter-intuitive and contests both the existing theories and intuitive beliefs we hold about emotion: Could there be issues with the reliability of ESM emotion time series that obscure a meaningful relation?

Because these issues with the internal and external validity of affect dynamics will be an important concern for next-generation ESM research to address (e.g., Wendt et al., 2020), this chapter provides an overview of different methodological strategies to increase the data quality of emotion time series. Essentially, each strategy fits in an overarching framework that has the ultimate goal to increase the signal-to-noise ratio (SNR) of the emotional time series ESM researchers investigate, and combining multiple strategies in your future ESM protocol will likely be most effective to reliably evaluate the internal and external validity of real-life affect dynamics (Dejonckheere et al., 2020).[1]

7.2 What Is the Signal-to-Noise Ratio of a Time Series?

While the SNR is an established metric in many other time series disciplines (e.g., audiology, Shojaei et al., 2016; biochemistry, Beal, 2015; econometrics, Saothayanun & Thangjai, 2018), it is a relatively unknown concept in the study of psychology. Although the operationalization and substantive interpretation may slightly vary across disciplines, the SNR can be easily applied to psychological time series data as well.

At its core, the SNR can be understood as a measure of data quality (Yu et al., 2018). As the name gives away, the SNR compares the power of desired signal to the power of background noise generated by a system (e.g., fMRI scanners, ecosystems, stock markets, etc.; Welvaert & Rosseel, 2013):

$$SNR = \frac{power\ signal}{power\ noise} \tag{7.1}$$

Broadly defined, the *signal* concerns the stream of meaningful information the system produces, while *background noise* refers to random (i.e., unpredictable) environmental interferences that distort that signal. As Eq. (7.1) illustrates, higher SNR values indicate higher data quality. Specifically, ratios higher than 1 imply that

[1] In this book chapter, we elaborate on some of the ideas formulated in our response *Reply to: Context matters for affective* chronometry (Dejonckheere et al., 2020) to Lapate and Heller's (2020) commentary on our original article (Dejonckheere et al., 2019a). Because this chapter constitutes a conceptual extension of our reply, some theoretical overlap is inevitable.

the power of a system's signal effectively outweighs the power of background noise, and that meaningful information can be detected. In contrast, SNRs smaller than 1 indicate that the power of a system's signal is exceedingly overruled by the power of random environmental influences, making it very hard to recover meaningful system information.

In the psychology of affect, researchers consider *emotional* systems (e.g., human beings responding to their natural environment; Loossens et al., 2020). Here too, the observed affective time series of ESM participants can be decomposed into a signal and noise term (Dejonckheere et al., 2020). A common and popular way to statistically operationalize the power of people's true emotional signal is to define it as the variance of a latent auto-regressive (AR) model of order 1 [i.e., an AR(1) model] (Schuurman et al., 2015; Schuurman & Hamaker, 2019):

$$power\ signal = \mathrm{Var}\left(\tilde{NA}_t = a\,\tilde{NA}_{t-1} + \varepsilon_t \right) \tag{7.2}$$

As can be seen from Eq. (7.2), an individual's true or latent (negative) affective signal $\tilde{NA}$ is defined by two terms. First, it is determined by a person-specific AR parameter (a) that captures the degree with which an individual's current latent affective state (e.g., a participant's true level of momentary $\tilde{NA}$ assessed at time point t) can be predicted from his or her previous latent affective state (e.g., a participant's true level of momentary $\tilde{NA}$ assessed at time point $t-1$). In dynamical system terms, this part of the equation refers to a person's *attractor strength* (Kuppens et al., 2010b), and reflects the speed with which an individual generally returns to his or her emotional baseline or homebase after responding to a stimulus that is of personal relevance (with higher a indicating slower return; Chow et al., 2005). Indeed, people's emotional reactions do not unfold in a vacuum, but are embedded in a larger context of personally relevant stimuli and events they respond to (Frijda, 1988). The collection of these contextual stimuli is situation-specific (i.e., different for each t), and cannot be captured by the person-specific AR parameter. They are therefore represented by a second term ε_t, generally referred to as *innovation* or *dynamic noise* (Schuurman et al., 2015). This dynamic noise term roughly corresponds to the intensity of the emotional stimulus a person reacts to at time point t, and carries over to next time points via the person-specific AR relation (hence the name *dynamic* noise; Schuurman & Hamaker, 2019).

Unfortunately, we are unable to directly observe people's true emotional signal. There is a degree of measurement error (ω) associated with each emotional assessment t that conceals participants' true momentary affective response. These random and unpredictable distortions are caused by moment-specific biases such as, for example, inattention, reactivity and fatigue (e.g., Fuller-Tyszkiewicz et al., 2013). Compared to innovation or dynamic noise, these interferences are restricted to each particular emotional assessment, and therefore do not resonate to subsequent assessments (Schuurman et al., 2015; Schuurman & Hamaker, 2019). The variance in measurement error across all momentary assessments represents the power of *measurement noise*:

$$power\ noise = \mathrm{Var}\left(\omega_t\right) \tag{7.3}$$

In sum, the combination of Eqs. (7.2) and (7.3) illustrates that participants' observed level of momentary (e.g., negative) affect can be understood as:

$$NA_t = a\,\tilde{NA}_{t-1} + \varepsilon_t + \omega_t \tag{7.4}$$

Accordingly, we can define the SNR of an affective time series as:

$$SNR = \frac{\mathrm{Var}\left(\tilde{NA}_t = a\,\tilde{NA}_{t-1} + \varepsilon_t\right)}{\mathrm{Var}\left(\omega_t\right)} \tag{7.5}$$

7.3 Determinants of the Signal-to-Noise Ratio

Equation (7.5) explains us how we can impact, and ultimately maximize the SNR of the affective time series we investigate. After all, attempts to advance the SNR lead to better data quality in ESM (Yu et al., 2018), promoting a more reliable evaluation of the internal and external validity of real-life affect dynamics.

But how troublesome is the SNR situation in contemporary ESM research really? To get an indication, we reanalyzed 15 prototypical ESM studies reported in Dejonckheere et al. (2019a), and computed the SNR for participants' PA and NA time series following Eq. (7.5) (see also Dejonckheere et al., 2020). As shown in Fig. 7.1, many of the datasets had median SNRs that barely exceeded the critical threshold of 1, which would indicate that 50% of the participants in that particular ESM study presented affective time series in which the level of estimated measurement noise overruled the power of their latent emotional signal. For the total sample in our meta-analysis (n = 1777), 30% of the subjects had an emotional SNR smaller than 1 for PA, and 42% for NA. Although each individual ESM study was carried out to answer a different research question, and their protocols, affect items and sample characteristics are therefore unique in design, these numbers suggest that the quality of current ESM time series is below par. This could explain why real-life affect dynamics have poor internal and external validity (e.g., Bulteel et al., 2018; Dejonckheere et al., 2020; Wendt et al., 2020): Current practices in ESM research prohibit the detection of meaningful dynamical regularities in emotion time series.

Then how may we boost the SNR of ESM time series? To guide the reader through the next sections, Fig. 7.2 presents a graphical overview of the different determinants that make up this metric (a, ε and ω), which researchers can impact to improve the SNR. For each determinant, we visualize two simulated affective time series that result in a low and high SNR (while the value of other determinants is

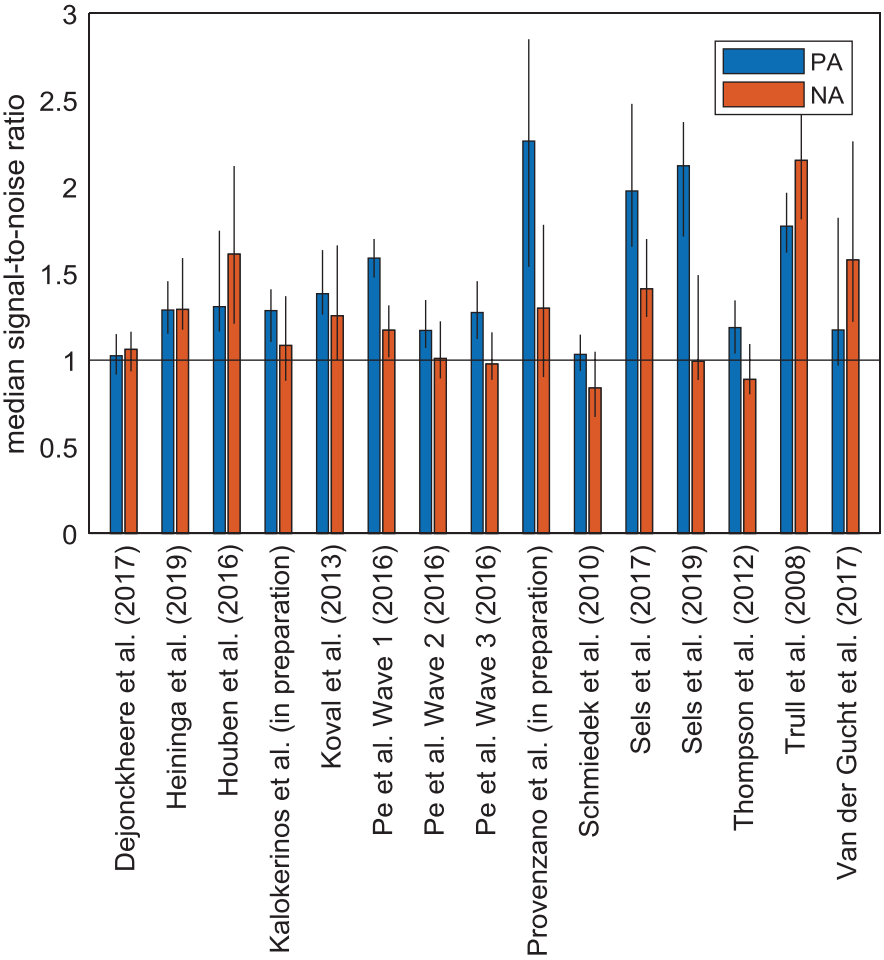


Fig. 7.1 Evaluating the SNR in conventional ESM research. The median SNR for PA (*blue bars*) and NA (*red bars*) for each of the 15 datasets reported in Dejonckheere et al. (2019a) (n = 1777), with the error bars indicating the 95% confidence interval (derived from 2000 bootstraps)

held constant). Metaphorically, you can think of the time series depicted in Fig. 7.2 as a pearl necklace: The string reflects a person's latent continuous emotional response ($\dot{NA}$), while the beads represent a series of discrete affect ratings (NA_t). To increase the SNR of an emotional time series, researchers should pursue substantially pronounced emotional strings (ε), and place the individual beads in such a way that they closely mirror the participant's original emotional response (both in time [a] and in deviation [ω]). In the next paragraphs, we will discuss how this analogy translates into concrete guidelines for each determinant.

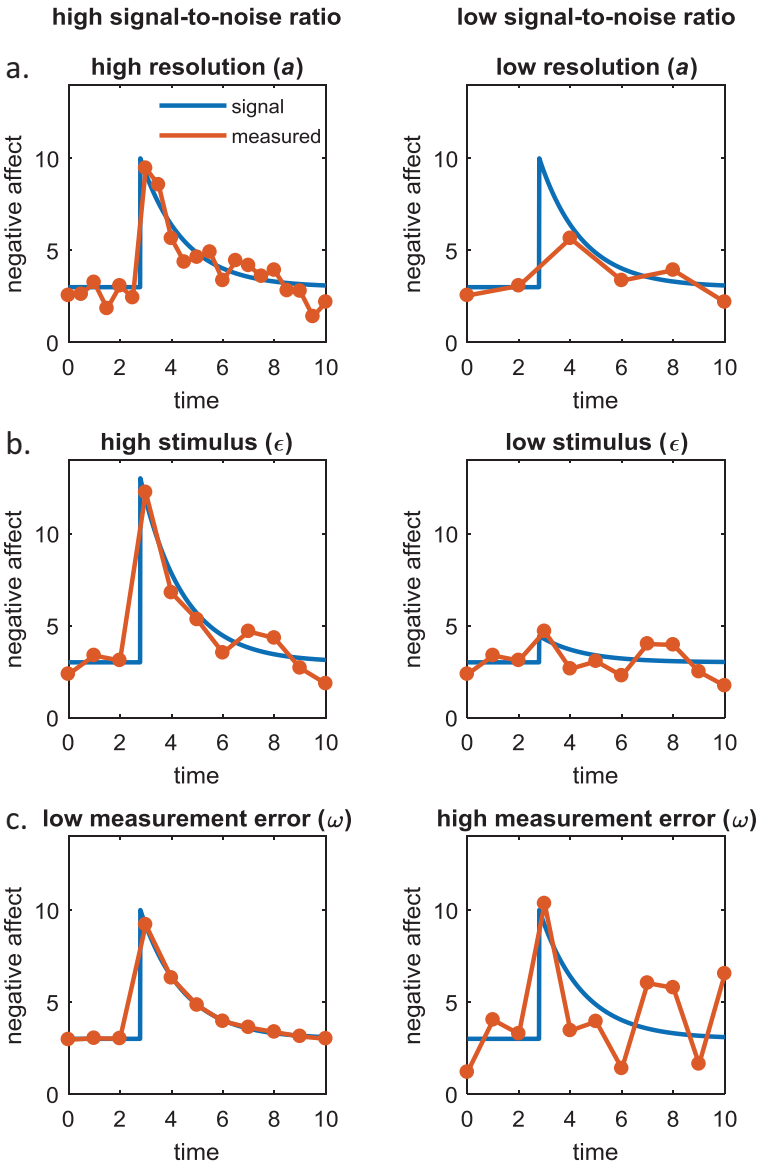


Fig. 7.2 Graphical representation of the various constituents (a, ε and ω) that make up the SNR. Hypothetical emotional trajectory of an individual who reacts to a negative contextual stimulus at time point 3. The *continuous blue line* represents that person's actual (latent) emotional response, while the *discrete red dots* depict his (observed) affective intensity ratings. The left and right columns refer to scenarios that exhibit a high and low SNR, respectively. (**a**) Manipulation of the temporal measurement resolution (short versus long time intervals). (**b**) Manipulation of the stimulus' emotional intensity (strong versus low emotional intensity). (**c**) Manipulation of the measurement noise associated with the affect ratings (assessments that are noise-free versus overwhelmed by measurement noise)

7.3.1 Recovering the Latent AR Parameter: Temporal Measurement Resolution

As mentioned earlier, the AR parameter *a* captures the degree with which a person's latent emotional reaction generally lingers after an affective stimulus was introduced (Kuppens et al., 2010a), and indicates how well we can predict its future (linear) trajectory (Headey & Wearing, 1989). When *a* is high, we describe a person's emotional recovery function as highly inert and resistant to the influence of internal regulatory processes (i.e., low attractor strength; see Fig. 7.3 panel A; Kuppens et al., 2010b). In contrast, when *a* is low, that person's emotional responding is highly susceptible to the system's regulatory forces, which will impact its original trajectory (i.e., high attractor strength; see Fig. 7.3 panel B).

Following this interpretation, *a* can be understood as a person-level indicator of the (linear) self-predictability of an affective state over time, and people's (observed) affective time series are known to vary in the amount of auto-regression they exhibit (Bos et al., 2019; Bulteel et al., 2018). Although *a* is considered person-specific, and we therefore cannot directly boost its magnitude to maximize the SNR, the upper limit of its estimation is determined by measurement practices (Dejonckheere et al., 2020). Specifically, because *a* refers to the temporal (un)predictability of a latent affective state (i.e., *How well can I predict a person's true emotion level at the next time point?*), the relative size of this parameter will naturally depend on the temporal assessment resolution that was used during the study protocol (see Fig. 7.2 panel A): Larger time intervals between consecutive measurement occasions t–1 and t will diminish the predictive accuracy of future affective states, and typically result in lower AR parameters (Bulteel et al., 2018). For example, making the average time interval between two consecutive measurement twice as long will reduce the estimated AR parameter by a power of 2 (e.g., from .40 to .16).

Intuitively, this makes sense, as the lingering effect of an emotional response generally diminishes as time after the instigating stimulus elapses (Anderson & Adolphs, 2014; Hemenover, 2003). As such, a measurement protocol in which successive measurements are simply too far apart will not be able to adequately capture the meaningful recovery of a person's emotional system, because full recovery likely took place in between measurement occasions (e.g., Schiepek et al., 2016). In other words, as a guiding principle, it is crucial that the temporal spacing between consecutive measurements is smaller than the rate of change of the emotions under study (Boker et al., 2009). When researchers assess emotional change with a higher temporal resolution, the serial dependency between consecutive emotional assessments increases (Ram et al., 2017), which allows for more fine-grained distinctions between individuals in terms of their personal AR parameter.

To illustrate the impact of a study's adopted assessment resolution on the recovery of participants' individual AR parameter, Fig. 7.4 visualizes three empirical ESM protocols from previous research that significantly differ in the average time interval between their emotional assessments (i.e., 1 day versus 104 min versus 13 min, respectively). For each protocol, we present the distribution of personal AR

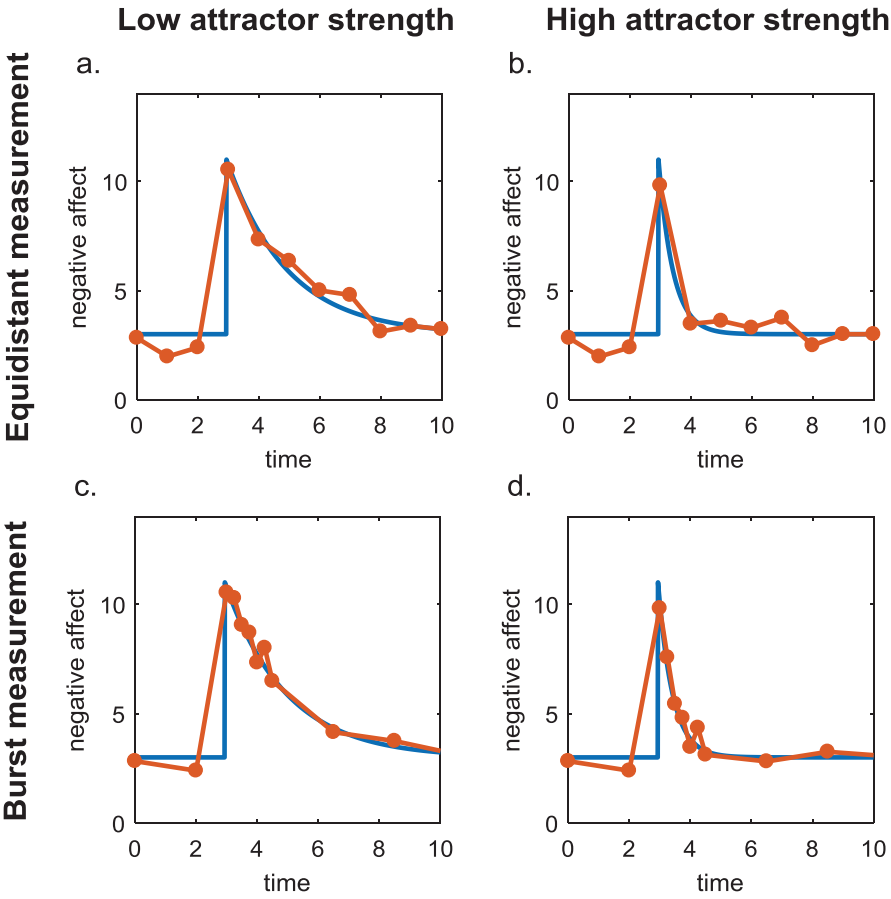


Fig. 7.3 Comparing the AR estimation in an equidistant measurement protocol versus micro-level measurement burst design. Simulated emotion time series for different hypothetical participants who react to a negative contextual stimulus at time point 3. The *continuous blue lines* represent their actual (latent) emotional response, while the *discrete red dots* depict their (observed) affective intensity ratings. (**a**) Participant who exhibits low attractor strength in an equidistant measurement protocol, which is adequately captured by the AR estimation. (**b**) Participant who exhibits high attractor strength in an equidistant measurement protocol, which is poorly captured by the AR estimation. (**c**) Participant who exhibits low attractor strength in a micro-level measurement burst design, which is adequately captured by the AR estimation. (**d**) Participant who exhibits high attractor strength in a micro-level measurement burst design, which is adequately captured by the AR estimation

parameters for PA, alongside the corresponding boxplot. Although a direct comparison between protocols is difficult due to other study-specific characteristics (e.g., total study duration, number of items per assessment, unique participant features, etc.), the boxplots suggest that longer time windows result in lower emotional AR estimations. Moreover, when longer time intervals between assessments exist, the proportion of participants that shows an AR parameter that does not significantly

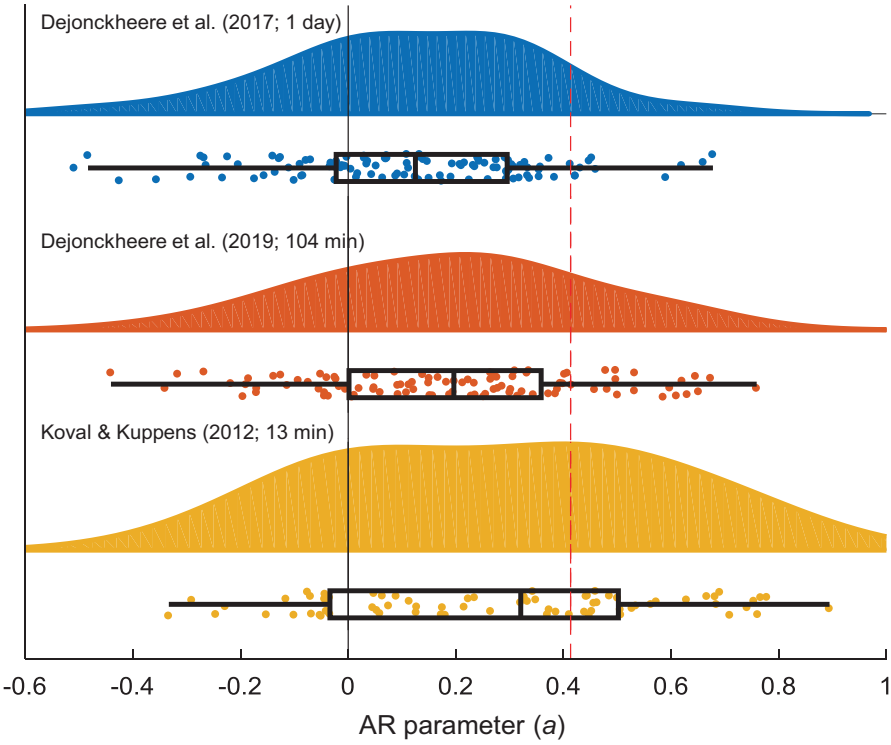


Fig. 7.4 Evaluating the role of temporal assessment resolution on the estimation of people's AR parameter. For each dataset, the raincloud plot visualizes the distribution of participants' personal auto-regressive effects in PA (i.e., the mean of emotion items *relaxed* and *happy*). The average time interval between two measurement occasions is presented between brackets. To maximize a fair comparison between studies, we only selected the first 30 affect ratings for each participant to estimate *a*. The *red dashed line* indicates the corresponding significance threshold

differ from zero increases (for an estimation based on 30 time points; i.e., 90, 82, and 62%, respectively). Conceptually, a non-significant AR parameter seems counter-intuitive, as it suggests that the momentary affect ratings of such an individual are completely instantaneously determined, and no carry-over effect in the form of emotional recovery between assessments takes place (i.e., there is no *emotional* inertia). Because such a motion is highly unlikely in real life, we assume that the temporal resolution of the assessment protocol was simply too low to adequately capture these individuals' emotional recovery.

But how do we determine the appropriate time scale with which emotions should be assessed? Ideally, a scenario in which we continuously monitor changes in an affective state (i.e., where the interval between t–1 and t is virtually zero) would enable the most accurate AR estimation. Following our necklace metaphor, this would imply that ESM researchers drastically increase the number of individual beads, in order to adequately reconstruct participants' underlying emotional

recovery function. However, to date, we can only access changes in experiential phenomena via repeated discrete self-reports, which is inherently intrusive and burdensome for participants (Schimmack, 2003; Stone et al., 2003). As such, these constraints preclude an endless expansion of the assessment frequency adopted in an ESM protocol.

Alternatively, instead of focusing on the appropriate *number* of assessments, ESM researchers should be equally concerned with the *timing* of these assessments (i.e., the exact temporal location of the different discrete beads to recover people's latent emotional signal; Kuiper & Ryan, 2018). That is, although inflating the number of discrete measurement occasions will result in shorter time intervals between occasions (when the study duration is held constant), chances still exist these assessments largely fail to accurately describe people's emotional recovery function when they are prompted at wrong moments (e.g., Fig. 7.3 panel B). Instead of holding on to measurement intervals that are more or less equidistant in time (which is common practice in contemporary ESM research), a more tailored ESM protocol that temporarily amplifies the rate of affective assessments when participants significantly deviate from their affective homebase, may enable a more accurate evaluation of their personal AR parameter (e.g., Schiepek et al., 2016). The flexible interval width of this *micro-level measurement burst design* (e.g., Stawski et al., 2015) softens the natural trade-off between increasing sample frequency and participant burden: Researchers may adequately assess participants' emotional recovery function while the total number of to be completed assessments should not be drastically increased (e.g., compare upper versus lower panels in Fig. 7.3). Although this approach may enable a better AR estimation, possible impediments to this type of measurement design include the implementation of online computational models in ESM software to determine whether participants are out of their personal affective equilibrium, the acquaintance with statistical models that do not require equally spaced time points (e.g., continuous-time Ornstein-Uhlenbeck models; Oravecz et al., 2009), and variable study durations across participants (when total number of assessments is held constant) that could lead to differences in the ecological validity of people's affective time series.

7.3.2 *Maximizing the Event-Specific Noise Term: Strong Contextual Stimuli*

Innovation or dynamical noise ε refers to the variance in a person's affective responding that cannot be explained by the latent AR model (Schuurman et al., 2015). Consequently, this error term is thought to capture change in people's emotional trajectory that does not stem from endogenous feedback processes (i.e., the person-specific AR parameter), but from the various contextual stimuli people encounter and react to in their environment. Although this noise term ε is serially uncorrelated and specific for each t, it may shape the value of subsequent latent

emotional states via the AR relation (Dejonckheere et al., 2020). As such, this situation-specific error term ε more or less refers to the (subjective) emotional intensity of the stimuli that people process.[2] For example, the sudden suffering of a goal of one's favorite soccer team will trigger an emotional reaction that cannot be accurately predicted from a person's previous emotional state (because it is unexpected), but it will have its effect on subsequent emotional assessments (in the form of emotional recovery).

In line with our necklace analogy, ε reflects the shape of people's underlying emotional string (i.e., explosive versus flat). Strong affect-laden stimuli, situations or events evoke strong emotional reactions, and increase the likelihood that people are temporarily thrusted out of their affective homebase (Chow et al., 2005). In turn, the emotional recovery process will be sizeable, which evidently produces a more pronounced emotional signal (see Fig. 7.2 panel B). The more pronounced this emotional signal, the easier it is to detect, as it effectively overrides the inevitable background noise associated with measurement practices. As such, exposing participants to events or situations that elicit strong emotional reactions may offer new opportunities to increase the SNR of the respective emotional ESM time series (Dejonckheere et al., 2020; Lapate & Heller, 2020).

At first sight, it may feel counter-intuitive that ESM researchers would want to control the contextual input their participants receive. After all, this data collection method is recognized for its ecological validity (Myin-Germeys et al., 2018; Trull & Ebner-Priemer, 2009), and is used to understand how emotions naturally come and go in the realm of ordinary life without any top-down interventions. This is in stark contrast with traditional lab research, where researchers experimentally induce isolated emotional states under controlled and standardized circumstances, and therefore have perfect insight in the causes and consequences of an emotional response. Because of the intrinsic tension between ecological validity and standardization, a downside to ESM (compared to experimental studies) is that real-life affect ratings are often equivocal, determined by a complex synergy of multiple ill-defined stimuli (Dejonckheere et al., 2019a). Not only are ESM researchers relatively clueless about the exact emotion-eliciting stimuli that shape participants' affective responses, the situations and events people experience on a daily basis are often emotionally unprovocative and short-lived (Dejonckheere et al., 2020). Bypassing these inherent limitations of ESM to get a closer indication of the contextual input participants receive, without carelessly undermining the ecological validity of this method, may boost the emotional signal value found in ESM time series.

To counter the fact that emotion ratings in daily life are often multi-determined, and that it is therefore challenging to isolate their direct cause, anchoring affective

[2] Some frameworks in the affect dynamics literature (e.g., Loossens et al., 2020) additionally break down ε_i into an innovation part (that captures deterministic contextual input) and a stochastic part (that captures built-in system noise). A more detailed discussion of this subdivision is beyond the scope of this chapter, but it explains the fractional translation of ε into the (perceived) emotional intensity of a stimulus.

assessments to specific events or stimuli could be worthwhile. In anchored ESM protocols, researchers track emotional change in daily life in relation to particular incidents or affairs. Instead of global momentary emotional assessments (e.g., *How sad do you feel right now?*), participants are instructed to rate their momentary emotion levels with regard to a specific and well-defined stimulus (e.g., *How sad do you feel right now, regarding your* [*partner, back pain, pregnancy, exams, job, etc.*]?). The selection of relevant stimuli may be fixed and determined a priori when researchers have a specific research question in mind (e.g., How do different emotions in the context of recent parenthood unfold?: *How* [*stressed, happy, etc.*] *are you about your new-born right now?*) or could be introduced bottom-up in the moment by the participants themselves. Here, participants first report on the event or stimulus that is currently most emotionally relevant via an open-ended question (e.g., *Describe the event or affair that triggers your current emotions.*). Next, via follow-up assessments, the natural emotional recovery with respect to the articulated event is evaluated, and new momentary relevant stimuli are disclosed that will serve as an input for subsequent affective assessments. Although this design would allow ESM researchers to delineate ecological emotional responses in relation to isolated stimuli, the possibility to assign conditional dependencies across measurements (where participants' responses at time t serve as input for their evaluations at time $t + 1$, $t + 2$, etc.) is currently lacking in many contemporary ESM applications. Furthermore, it remains to be investigated to what extent people are truly able to detach their emotional evaluation related to a single stimulus from their global affective state (e.g., carry-over effects between stimuli, Schmidt & Schmidt, 2016), and if these type of assessments do not induce stronger emotional reactivity (e.g., Schwarz, 1999).

Anchoring emotional assessments in ESM does not tackle the problem that the events and situations people encounter on a regular basis may not be very emotionally moving (i.e., their emotional string is not very pronounced). Moreover, because individuals act in their own personal ecology, the contextual input they receive may differ in numerous ways (Kahneman et al., 2004), which hampers a direct comparison of their affective time series. Here, quasi-experimental ESM studies, in which researchers track natural emotional change around a standardized and impactful stressor could be promising (Dejonckheere et al., 2020). Researchers may introduce that stressor into the daily lives of their participants top-down (e.g., a Trier Social Stress Task; Koval & Kuppens, 2012), or they may anticipate the occurrence of a real-life event to design a study around (e.g., the release of exam results, Dejonckheere et al., 2019b; Kalokerinos et al., 2019; Metalsky et al., 1993; the implementation of COVID-19 lockdown measures, Dejonckheere et al., 2021; Taquet et al., 2020; the onset of a depressive episode, Wichers et al., 2016; the death of a beloved one, Folkman, 1997). In both cases, all subjects are exposed to the same compelling stimulus, which allows a more controlled comparison of their emotional trajectory, but real-life stressors have the strength that they are unparalleled in their ecological quality. However, a possible downside to these naturalistic stimuli is that they are sometimes relatively difficult to predict prospectively, which may result in longer study durations (e.g., Folkman, 1997) or smaller sample sizes

(e.g., Dejonckheere et al., 2021; Wichers et al., 2016). Nevertheless, this quasi-experimental approach can reconcile both the naturalistic qualities of ESM and the more standardized setting of lab experiments (Dejonckheere et al., 2019b).

7.3.3 *Reducing Momentary Measurement Noise: Assessing Measurement Error*

Finally, ESM researchers can also improve the SNR of an affective time series by diminishing the measurement error ω that is associated with their assessment procedures (see Fig. 7.2 panel C). Attempts to reduce the incidental distortions inherent to people's momentary affect ratings may produce a closer approximation of their true emotional response, and therefore minimize the bias present in their AR parameter estimation (e.g., Staudenmayer & Buonaccorsi, 2005). Thus, in line with our necklace metaphor, ω indicates the extent with which the individual beads properly link up with the underlying string, indicating how much the discrete affect ratings deviate from a person's latent emotional signal.

Although the ubiquity of measurement noise in affective time series is commonly accepted among ESM researchers, this issue is largely ignored in the affect dynamics literature (e.g., Schuurman et al., 2015; Schuurman & Hamaker, 2019). Not only do emotion researchers often calculate affect dynamic measures from observed affect intensity scores (Dejonckheere et al., 2019a; Wendt et al., 2020), in an attempt to minimize participant burden or annoyance, they are generally reluctant to include exact repetitions of an item within the same momentary questionnaire (Schimmack, 2003). Singular items, however, prohibit an explicit momentary reliability assessment, leaving researchers in the dark about the exact amount of measurement error present in people's affect ratings. Instead of indirectly evaluating measurement noise via estimation procedures, future ESM designs could therefore benefit from randomly repeating (at least) one item at each measurement occasion, in order to get a direct indication of the measurement error associated with their protocol.

In a derivative effort to somehow mitigate the bias of measurement noise in affective time series, it is common practice in ESM research to average an *ad hoc* selection of specific same-valenced emotion items to construct a global positive or negative affective composite (Dejonckheere et al., 2019a). Combining individual emotion items into a single construct is thought to compensate for some of the measurement error associated with each individual rating (e.g., Nunnally, 1994), which may produce a higher SNR for the aggregated time series. To evaluate how this averaging procedure impacts the SNR, we compared the SNRs of each individual emotion item versus the two global PA and NA composites for a traditional ESM dataset (Sels et al., 2017). As shown in Fig. 7.5, the SNR for the two affective aggregates PA and NA is remarkably higher than for those of the same-valenced individual emotion items (except for *stressed*).

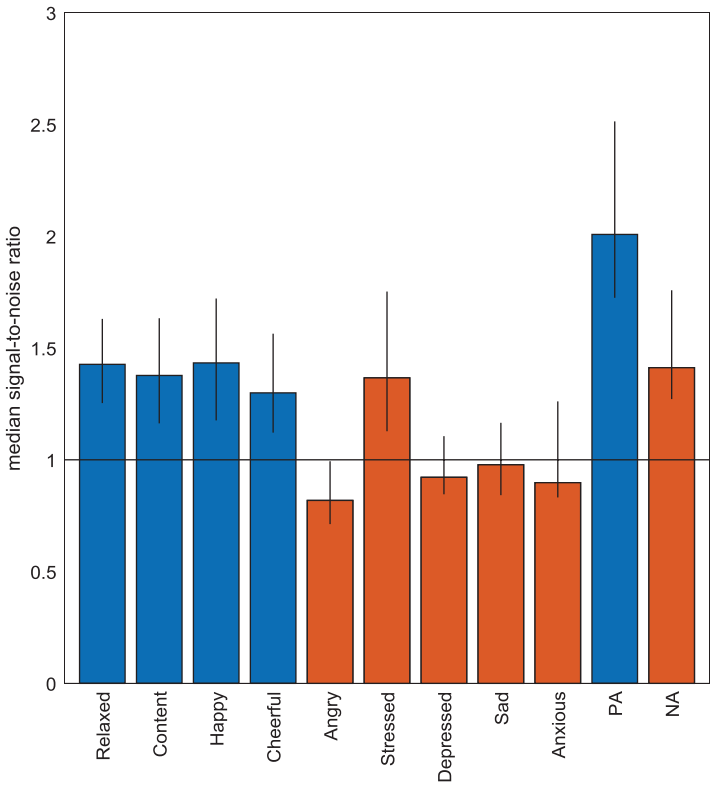


Fig. 7.5 Investigating the SNR of individual emotion items versus global PA and NA constructs in Sels et al. (2017). The median SNR for positive (*blue bars*) and negative (*red bars*) affect items or constructs, with the error bars indicating the 95% confidence interval (derived from 2000 bootstraps)

Nevertheless, when an emotional time series carries little signal, attempts to reduce measurement error will have little effect on the overall SNR. As such, procedures to reduce measurement noise are best used in combination with signal improvement strategies, because it is an illusion to believe researchers can entirely eliminate the measurement error associated with psychological self-report. Furthermore, it should be noted that the within-person internal consistencies of multi-item PA and NA constructs remain generally low (Dejonckheere et al., 2019a). This leaves the question to what extent the adopted discrete emotion items are truly interchangeable indicators of a global PA or NA composite, and what these averaged constructs really capture if their momentary reliabilities are so remarkably low. A more considerate and theory-driven selection of specific emotion-items could result in global PA and NA constructs that are both internally reliable and less sensitive to measurement error.

7.4 Combining Different Strategies to Improve the Signal-to-Noise Ratio

In this chapter, we introduced an overarching framework to promote the SNR of emotional time series in ESM research. Drawing from Eq. 7.5, this framework was structured around three key determinants that make up an emotional SNR. On the one hand, we discussed how researchers could seek to increase the power of participants' true emotional signal by fine-tuning AR estimations (a) or investigating more pronounced emotional reactions (ε). On the other hand, we advised researchers to evaluate and to reduce the measurement noise (ω) associated with their ESM protocol by creating reliable same-valenced affective composites instead of relying on singular emotion items. In the previous paragraphs, we explained how these general propositions translated into concrete design choices for each determinant individually (see Table 7.1 for an overview), but ideally ESM researchers may wish to apply these strategies to their protocol simultaneously, in order to optimize the quality of their time series data in the best possible way.

As a proof-of-concept that the combination of outlined strategies adds to an improved SNR, Panel A of Fig. 7.6 visualises how the median SNR of a real ESM study sample changes under different (artificial) strategy scenarios. In this ESM study (Dejonckheere et al., 2019b; Kalokerinos et al., 2019), we tracked the emotional trajectories of 101 first-year students around an impactful and personally relevant event, the release of their exam results. Students were instructed to rate both their unanchored momentary PA and NA (*Please indicate how positive/negative you are feeling right now?*), as well as multiple discrete emotion items anchored to their grades (*When you think about your grades right now, how* [*content, happy, proud, relieved, angry, anxious, ashamed, disappointed, stressed*] *are you feeling?*). Same-valenced emotion items were averaged at each measurement occasion to create an anchored PA and NA time series, and we computed an additional global anchored affective construct in which combined all items together (PA-NA). Finally, to simulate scenarios with different temporal resolutions, we relied both on participants' original time series, as well as a trimmed version in which we only considered every fifth emotional assessment.

Table 7.1 Summary of the proposed design strategies for each SNR determinant

SNR goal	Design strategy
Improve AR estimation (a)	– Increase temporal assessment resolution – Micro-level measurement burst design
Boost situational input (ε)	– Anchor emotional assessments to specific stimuli – Introduce/anticipate strong contextual events
Reduce measurement noise (ω)	– Explicitly assess measurement error via exact item repetitions – Reliably combine same-construct items

SNR Signal-to-noise ratio, *AR* Auto-regression

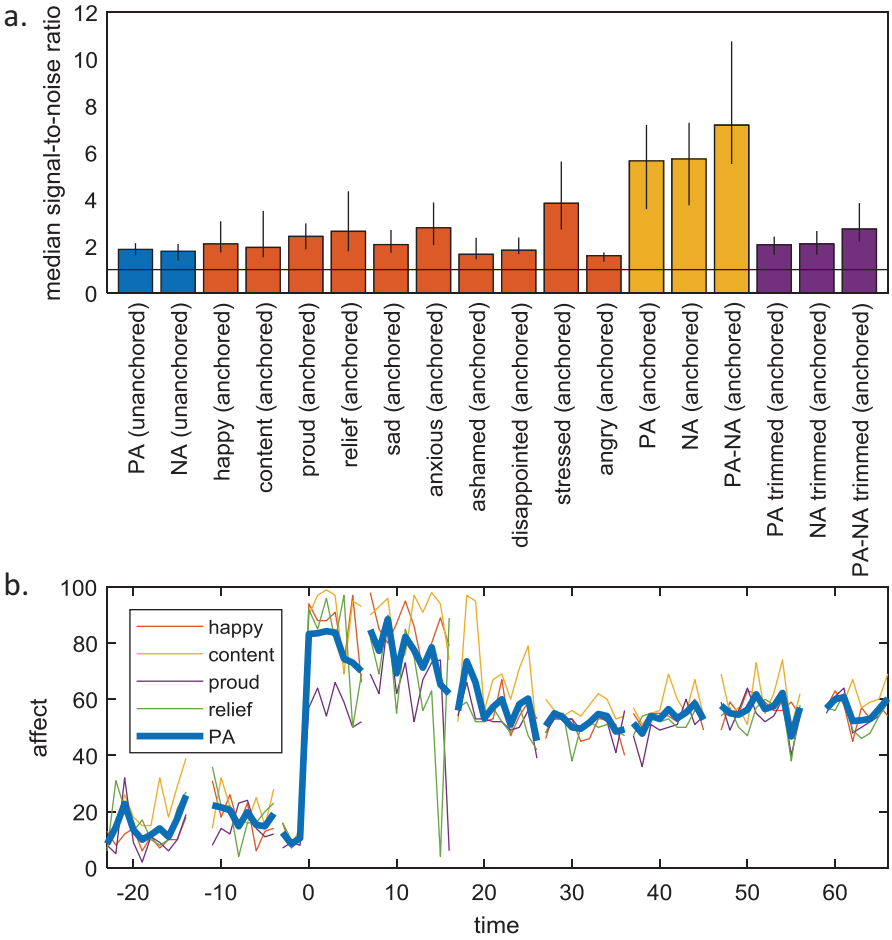


Fig. 7.6 Combining multiple strategies to improve the SNR in ESM time series. The results in both panels rely on data reported in Dejonckheere et al. (2019b) and Kalokerinos et al. (2019), in which we followed the emotional trajectories of 101 first-year students around the time they received their exam results. (**a**) The median SNR for different strategy scenarios, with the error bars indicating the 95% confidence interval (derived from 2000 bootstraps). (**b**) Real affective time series for an example participant with a high SNR (22.43). Time point zero indicates the first emotional assessment after the student consulted his or her exam results

As Panel A of Fig. 7.6 suggests, implementing multiple strategies in an ESM protocol at once markedly improves the SNR of emotion time series. First, for each scenario, the median SNR is almost around twice as high than those of most of the traditional ESM studies in Fig. 7.1, hinting at a positive impact of studying strong contextual stimuli on participants' emotional signal. Second, a comparison of the unanchored PA and NA items versus the anchored assessment of different discrete emotions shows that some (but not all) anchored emotion items bring about slight increases in the SNR (e.g., *stressed* but not *angry*). This suggests that assessing

(some) emotional states in relation to a specific stimulus could potentially benefit the SNR. Third, averaging single anchored emotion items into an anchored global PA and NA composite drastically boosts the SNR, and its value increases even more when a global affective composite is considered (PA-NA). This indicates that the practice of averaging affect items reduces the measurement error associated with each individual emotion rating. Finally, when comparing the anchored PA and time series of the trimmed versus complete dataset, the SNR is considerably higher when a more fine-grained temporal assessment resolution is adopted. This suggests that an AR estimation is more accurate when the time interval between consecutive measurements is compressed.

But how does a high SNR visually manifest in an empirical ESM time series? Panel B of Fig. 7.6 depicts the PA time series for a participant with one of the highest SNRs in the study sample (SNR 22.43). First, the unexpected and sudden jump around the release of that participant's exam results indicates the introduction of a strong emotional stimulus, kicking that person out of emotional equilibrium and allowing emotional recovery to take place. Second, the aggregation of discrete emotion items into a global affective composite score clearly smooths the affective signal, eliminating some of the incidental and irregular drops and spikes that shape individual emotion ratings (which may be attributed to measurement error).

7.4.1 Interdependencies Among Design Strategies

Although the results in Fig. 7.6 suggest that implementing multiple design strategies positively amplifies the SNR, it is important to acknowledge that their effect is not necessarily additive. Similarly, the separate review for each individual SNR determinant does not imply that each design strategy independently impacts the SNR. As such, mutually comparing the effect of different design strategies is probably meaningless. There may be positive structural dependencies between the different strategies we discussed, making it difficult to disentangle their unique contribution in improving the SNR.

In contrast, it is equally possible that negative associations between particular design strategies exist, carrying an opposite impact on the SNR. That is, a proposed strategy to improve one SNR determinant may unintentionally compromise another one. For example, repeatedly exposing participants to micro-level measurement burst cycles has the goal to improve AR estimations, but could also induce increased annoyance with the protocol, resulting in more measurement error. Similarly, investigating real-life emotions in relation to a personally relevant and impactful event may boost the innovation parameter, but could equally introduce more missing data due to the study's increased interference with people's lives, impeding accurate AR estimations. As a final example, multiple items per construct may reduce the measurement error associated with each individual question, but result in longer momentary assessments, which is known to predict poor compliance (Eisele et al., 2020).

Depending on how all of these design choices relatively impact each determinant, SNR values may increase, decrease or remain unaltered. In either case, this uncertainty calls for future ESM studies that explicitly test how the SNR changes in function of various design alternatives.

7.4.2 Design Strategy Implementation Constraints

Finally, we realize that many of the outlined propositions may currently be difficult to implement, and that the resulting ESM protocols drastically differ from conventional ESM research practices today. For one, there are practical constraints. For example, regarding the implementation of micro-level measurement burst cycles, the possibility to model people's emotional trajectory online (needed to instantaneously detect abrupt changes in affect), is currently lacking in standard ESM applications. Similarly, in the context of anchoring idiosyncratic emotional assessments, installing input-output dependencies between consecutive measurements is not straightforward with modern ESM software. Technical advances are needed to remove these barriers.

Second, some of the design strategies presented challenge the way ESM researchers traditionally model affect dynamics. For example, tracking people's emotional reaction in response to an impactful stressor likely yields time series that are not stationary, violating a statistical assumption that underlies some of the commonly investigated affect dynamics (e.g., emotional inertia or network density; Bringmann et al., 2013; Pe et al., 2015). Relatedly, the repeated use of measurement burst cycles violates the assumption of equally spaced time points, preventing for instance a standard assessment of people's global level of emotional instability (Jahng et al., 2008). In sum, potential adjustments to traditional ESM designs will close the door for some commonly studied affect dynamic metrics. At the same time, however, novel design strategies allow researchers to model dynamical patterns in affect in a more nuanced and fine-grained manner.

7.5 Conclusion

When interested in the real-life dynamics of emotion, this book chapter invites ESM researchers to raise the bar when it comes to the data quality of their studies. The SNR in traditional ESM research is typically substandard, which demands future daily life studies to experiment with more exotic design approaches to effectively disentangle people's true emotional reactions form inevitable background noise. Only then will we be able to reliably assess the internal and external validity of real-life affect dynamics.

References

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell, 157*, 187–200.

Beal, J. (2015). Signal-to-noise ratio measures efficacy of biological computing devices and circuits. *Frontiers in Bioengineering and Biotechnology, 3*, 93.

Boker, S. M., Molenaar, P. C. M., & Nesselroade, J. R. (2009). Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. *Psychology and Aging, 24*, 858–862.

Bos, E. H., de Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia–instability paradox. *British Journal of Psychology, 110*, 814–827.

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment, 23*, 425–435.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One, 8*, e60188.

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods, 23*, 740–756.

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion, 5*, 208–225.

Cole, P. M., & Hollenstein, T. (2018). *Emotion regulation: A matter of time*. Routledge.

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., Brose, A., Bastian, B., & Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology, 114*, 323–341.

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019a). Complex affect dynamic measures add limited information to the prediction of psychological well-being. *Nature Human Behaviour, 3*, 478–491.

Dejonckheere, E., Mestdagh, M., Kuppens, P., & Tuerlinckx, F. (2020). Reply to: Context matters for affective chronometry. *Nature Human Behaviour, 4*, 690–693.

Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., & Kalokerinos, E. K. (2019b). The relation between positive and negative affect becomes more negative in response to personally relevant events. *Emotion, 21*(2), 326–336.

Dejonckheere, E., Houben M., Schat, E., Ceulemans, E., & Kuppens, P. (2021). The short-term psychological impact of the COVID-19 pandemic in psychiatric patients: Evidence for differential emotion and symptom trajectories in Belgium. *Psychologica Belgica, 60*, 1–10.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*.

Folkman, S. (1997). Positive psychological states and coping with severe stress. *Social Science and Medicine, 45*, 1207–1221.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43*, 349–358.

Fuller-Tyszkiewicz, M., Skouteris, H., Richardson, B., Blore, J., Holmes, M., & Mills, J. (2013). Does the burden of the experience sampling method undermine data quality in state body image research? *Body Image, 10*, 607–613.

Headey, B., & Wearing, A. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology, 57*, 731–739.

Hemenover, S. H. (2003). Individual differences in rate of affect change: Studies in affective chronometry. *Journal of Personality and Social Psychology, 85*, 121–131.

Hisler, G. C., Krizan, Z., DeHart, T., & Wright, A. G. C. (2020). Neuroticism as the intensity, reactivity, and variability in day-to-day affect. *Journal of Research in Personality, 87*, 103964.

Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141*, 901–930.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13*, 354–375.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*, 1776–1780.

Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., Thompson, R. J., Van Ryckeghem, D. M. L., Kuppens, P., & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences, 117*, 9270–9276.

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychological Science, 30*, 863–879.

Kalokerinos, E. K., Tamir, M., & Kuppens, P. (2017). Instrumental motives in negative emotion regulation in daily life: Frequency, consistency, and predictors. *Emotion, 4*, 648–657.

Koval, P., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion, 15*, 625–636.

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion, 12*, 256–267.

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal, 25*, 809–823.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010a). Emotional inertia and psychological maladjustment. *Psychological Science, 21*, 984–991.

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010b). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*, 1042–1060.

Lapate, R. C., & Heller, A. S. (2020). Context matters for affective chronometry. *Nature Human Behaviour, 4*, 688–689.

Loossens, T., Mestdagh, M., Dejonckheere, E., Kuppens, P., Tuerlinckx, F., & Verdonck, S. (2020). The affective ising model: A computational account of human affect dynamics. *PLoS Computational Biology, 16*, e1007860.

MacCann, C., Erbas, Y., Dejonckheere, E., Minbashian, A., Kuppens, P., & Fayn, K. (2020). Emotional intelligence relates to emotions, emotion dynamics and emotion complexity: A meta-analysis and experience sampling study. *European Journal of Psychological Assessment, 36*, 460–470.

Metalsky, G. I., Joiner, T. E., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: A test of the hopelessness and self-esteem theories of depression. *Journal of Abnormal Psychology, 102*, 101–109.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*, 123–132.

Murray, G., Nicholas, C. L., Kleiman, J., Dwyer, R., Carrington, M. J., Allen, N. B., & Trinder, J. (2009). Nature's clocks and human mood: The circadian system modulates reward motivation. *Emotion, 9*, 705–716.

Nunnally, J. C. (1994). *Psychometric theory* (3rd ed.). Tata McGraw-Hill Education.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical ornstein–uhlenbeck model for continuous repeated measurement data. *Psychometrika, 74*, 395–418.

Pe, M., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science, 3*, 292–300.

Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. *Emotion, 12*, 1320–1328.

Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development, 14*, 253–270.

Robinson, M. D., Irvin, R. L., Persich, M. R., & Krishnakumar, S. (2020). Bipolar or independent? Relations between positive and negative affect vary by emotional intelligence. *Affective Science*.

Saothayanun, L., & Thangjai, W. (2018). Confidence intervals for the signal to noise ratio of two-parameter exponential distribution. In L. H. Anh, L. S. Dong, V. Kreinovich, & N. N. Thach (Eds.), *Econometrics for financial applications* (pp. 255–265). Springer International Publishing.

Schiepek, G., Aichhorn, W., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-time monitoring of psychotherapeutic processes: Concept and compliance. *Frontiers in Psychology, 7*, 604.

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies, 4*, 79–106.

Schmidt, S. R., & Schmidt, C. R. (2016). The emotional carryover effect in memory for words. *Memory, 24*, 916–938.

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods, 24*, 70–91.

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in N = 1 psychological autoregressive modeling. *Frontiers in Psychology, 6*, 1038.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93–105.

Sels, L., Ceulemans, E., & Kuppens, P. (2017). Partner-expected affect: How you feel now is predicted by how your partner thought you felt before. *Emotion, 17*, 1066–1077.

Shojaei, E., Ashayeri, H., Jafari, Z., Zarrin Dast, M. R., & Kamali, K. (2016). Effect of signal to noise ratio on the speech perception ability of older adults. *Medical Journal of the Islamic Republic of Iran, 30*, 342.

Staudenmayer, J., & Buonaccorsi, J. P. (2005). Measurement error in linear autoregressive models. *Journal of the American Statistical Association, 100*, 841–852.

Stawski, R. S., MacDonald, S. W. S., & Sliwinski, M. J. (2015). Measurement burst design. In *The encyclopedia of adulthood and aging* (pp. 1–5). American Cancer Society.

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain, 104*, 343–351.

Taquet, M., Quoidbach, J., Fried, E. I., & Goodwin, G. M. (2020). Mood homeostasis before and during the coronavirus disease 2019 (COVID-19) lockdown among students in The Netherlands. *JAMA Psychiatry.* https://doi.org/10.1001/jamapsychiatry.2020.2389

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment, 21*, 457–462.

Welvaert, M., & Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS One, 8*, e77089.

Wendt, L. P., Wright, A. G. C., Pilkonis, P. A., Woods, W. C., Denissen, J., Kühnel, A., & Zimmerman, J. (2020). Indicators of affect dynamics: Structure, test-retest reliability, and personality correlates. *European Journal of Personality*. https://doi.org/10.1002/per.2277

Wichers, M., Groot, P. C., & Psychosystems, E. G. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics, 85*, 114–116.

Yu, S., Dai, G., Wang, Z., Li, L., Wei, X., & Xie, Y. (2018). A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. *BMC Medical Imaging, 18*, 17.